

Extracting Touristic Information from Online Image Collections

Edoardo Ardizzone^{*}, Francesco Di Miceli^{**}, Marco La Cascia^{*} and Giuseppe Mazzola^{*}

Dipartimento di Ingegneria Chimica, Gestionale, Informatica, Meccanica
Università degli Studi di Palermo, viale delle Scienze bd. 6, 90128, Palermo, Italy

^{*}{edoardo.ardizzone, marco.lacascia, giuseppe.mazzola}@unipa.it
^{**}cicciodmc@gmail.com

Abstract— In this paper, we present a Geographical Information Retrieval system, which aims to automatically extract and analyze touristic information from photos of online image collections (in our case of study Flickr). Our system collect all the photos, and the related information, that are associated to a specific city. We then use Google Maps service to geolocate the retrieved photos, and finally we analyze geo-referenced data to obtain our goals: 1) determining and locating the most interesting places of the city, i.e. the most visited locations, and 2) reconstructing touristic routes of the users visiting the city. Information is filtered by using a set of constraints, which we apply to select only the users that reasonably are tourists visiting the city. Tests were performed on an Italian city, Palermo, that is rich in artistic and touristic attractions, but preliminary tests showed that our technique could successfully be applied to any city in the world with a reasonable number of touristic landmarks.

Keywords - Flickr, Geotagging, Tourism, Geographic Information Retrieval, GIS, Google Maps

I. INTRODUCTION

In recent years the growing integration between digital cameras and portable devices like mobile phones, PDAs or tablets, has dramatically increased the number of pictures taken from places all over the world. New Internet services (like Flickr, Panoramio, Smugmug, Picasa) were created to offer to the users online storage services for their photos, creating huge databases of photographs, that implicitly contain an immense amount of information. These online image collections consist not only of raw pictures, but also of tags, annotations, descriptions, and often information about geo-referencing. Each of these photos individually provides a low information content, but if we consider the huge amount of pictures of the whole online collections we are able to extract useful information for different purposes.

The aim of this work is to create a tool for touristic applications, using only the resources on the Web (photos, descriptions, text), without any explicit human intervention to organize this information. The key concept is that the content is already on the Web and we need just to retrieve it and organize it. In this way every single photo, tag or comment uploaded by a user on the Web become automatically a new information resource, and can be potentially exploited by any user who needs it.

In this work we decided to study the largest and most famous online collection of photos, Flickr. Our system automatically retrieves data (photos, tags, annotations) from this database and organizes them in a functional map locating the various points of interest (landmarks). Our system is able also to extract and reconstruct touristic routes of Flickr users, to study their activities when visiting a city. In section 2 we discuss about Geographic Information Retrieval systems, which dealt with problems that are similar to the one studied in this paper. In section 3 we present our information retrieval system. In section 4 we discuss about implementation and experimental results. A conclusive section ends the paper.

II. GEOGRAPHIC INFORMATION RETRIEVAL

The Geographic Information Retrieval (GIR) systems may be considered as a specialization of the typical Information Retrieval systems, that emphasizes the representation and retrieval of geographic content of documents [1]. The main features of this kind of systems are:

- Explicit indexing of geographic content
- Queries with spatial conditions and content

Geographic Information Retrieval does not deal only with objects that have a physical location in geographic place (as rivers, lakes, cities or countries), but refers to any type of information that has some relationships with one or more positions on the Earth's surface. This kind of information are known as geo-referenced information and these systems are also known as Georeferenced Information Retrieval systems. The analysis of geo-referenced information can be important in different contexts: for researchers, to view information about climate change in a certain area, or the population growth in a particular city; for more social purposes, such as finding a restaurant in a given area, or hotels, or shops; for touristic purposes to plan a route on the basis of spatial information.

A GIR system typically includes a text retrieval and processing module and a spatial information retrieval and processing module. We can identify the phases that characterize the execution of a query in a GIR system: textual query insertion, text processing, geospatial coordinates resolution, text-based IR, spatial matching, text matching, ranking, browsing and visualization.

The identification of geographic locations via text query implies some problems: names are not unique, as the name of a city for example is a fairly common name for a geographic location; changes in translation, as place names may differ depending on the language (Rome, Roma). Some systems in fact prefer to use directly pairs of textual and spatial queries.

The task of finding information often refers to the identification of geographical location or location-specific events. A classic application of a GIR system aims to create a map where the system must locate the places which can be of interest for a user. A growing number of methodologies have been developed to extract information from the Web in order to detect points of interest on city maps, mainly using image information. One reason is the presence of online collections, into which users every day download their photos from around the world. They represent a wealth of information, above all because most of these images have, in addition to visual information, metadata like tags and information about georeferencing. It gives the possibility to link information about pictures to a specific area, and to use it to give a semantics to a given point on a geographic map. Landmark detection methods exploit what is the so-called Location Semantics (or Place Semantics) [2] of the objects, i.e. the information that answers to the question "where the photo was taken?". In practice the Location Semantics make a connection between an object and a geographic entity, such as a city, or some more specific area, as a museum or a library. In the next section we will present some works related to Location Semantics.

A. State of the art

There is a growing interest on topics such as event and landmark detection, and automatic extraction of geographic information from large-scale social media contents, and it is evidenced by the large number of solutions proposed for various applications.

Many works studied the problem of organizing pictures related to the tourist attractions of a city. Among these [3] has proposed to derive, from thousands of photographs downloaded from Internet sharing sites, a summary page that consists of images that capture the most interesting sites of a city, pulling out a canonical image for each representative point. This approach extracts visual features from the images, by using the SURF algorithm, and uses a partitioned clustering, based on a k-means, to group similar photos. An interesting approach which uses image processing techniques to exploit online collections is [4], in which through the use of SIFT feature and image matching, the authors try to build a 3D reconstruction of Rome. This project required the processing of a huge amount of photographs, needed to place the graphic details of a 3D reconstruction. For this reason the implemented techniques are mainly related to graphical analysis and time optimization for data processing.

User tags can be exploited as an additional source for spatio-temporal discrimination between images, and they can enrich the landmark detection techniques. In [5] authors used positional clustering techniques and tag scoring to generate a set of keywords that describe the content of a collection of geo-referenced images, and use

these tags to mark an area on a map with a representative word. A similar approach was used by [6] which implemented a service, called World Explorer, that can allow users to navigate a map using representative tags. Another similar approach is [7] in which the authors developed a system which used a fusion of text, positional and visual processing techniques. In addition to positional clustering and TF-IDF metrics, used in both [5] and [6], in [7] an image classification based on the visual content is used to further group images that were already clustered using positional and textual information. A practical application of these techniques is given in [8], which describes a service for the retrieval of touristic images. The application first performs a positional clustering onto images from Flickr, to find some points of interest. Then it makes a scoring of the image tags of each cluster using the TF-IDF algorithm. Finally, for each cluster, it acquires new images by another query, based on the extracted tags, and make a visual matching for each point to derive the most representative image. The final result states that every point of interest is depicted on the map using the most representative image.

Other methods focus also on the event semantics, i.e. the anchoring of an object to a specific event. In [9] the proposed method divides an area of a map in portions, and for each portion analyzes the tags statistics, to give a meaning of place or event to the photo tags, studying the number of users and the photo bursts. Statistics on photo timing are also used in [10], in which the authors want to get a system that answers to the questions: "How long I'll take to visit this touristic attraction?" or "what can I visit in one day in this city?". In particular, the method uses the statistics of temporal metadata in geo-referenced images, trying to estimate the duration of the visit of a user in a specific location. In [11], in addition to geo-referenced information, the authors exploit a different type of metadata: the user interests. The system search for images on both a generic search engine and Flickr, and Flickr images are sorted according to the users preferences. Then images are clustered according to visual and textual features and the obtained clusters inherit the evaluation from the images they include. In the final step images previously downloaded by a query to a generic search engine are reordered according to the obtained clusters and presented as divided by the point of interest. In [12] the authors analyze the relationship between travel information, extracted from community-contributed photos, and user attributes (e.g., gender, age, race), to automatically suggest new travel routes to the users.

III. PROPOSED METHOD

In this paper we present a Geographic Information Retrieval system which aims to extract touristic information analyzing photos of Flickr users. The goal is twofold: finding places of interest in a city and reconstructing touristic routes of Flickr users that visited that city.

A. System Overview

We first collect from Flickr all the photos, and related information, that are associated to a specific city, by using

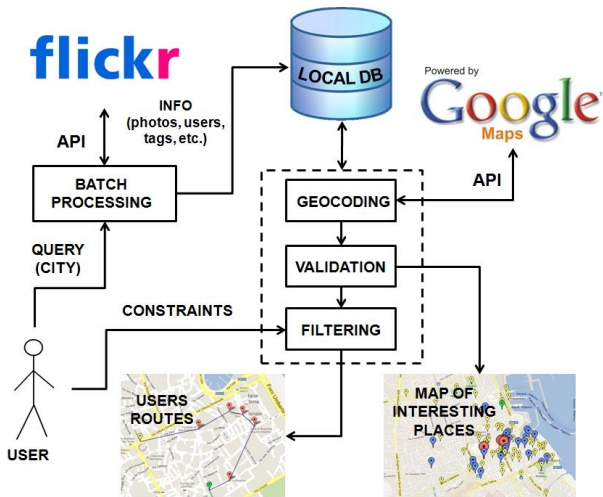


Figure 1. Scheme of the proposed system.

the Flickr API. For our purpose and to save execution time in the processing step, we stored the retrieved information into a local database. The most important information that can be retrieved by a query to Flickr are:

- **photo.exif**: in which are stored the main physical information of the individual photos (e.g. timestamps);
- **photo.note**: in which are stored all the notes related to a photo;
- **photo**: which stores the “social” information of the photo;
- **photo.comment**: where are stored all the comments on a photo;
- **photoset**: that stores information about set of photos;
- **photoset.comment**: which saves all the comments on a particular photo set;
- **place**: where are stored the GPS coordinates (if they exist) of where the photos are taken by users;
- **photo.tag**: that stores all the relevant tags of each photo;
- **user**: it stores information about users;
- **group**: in which are stored the information of the groups the users belong to;
- **contact**: in which all the user's contact info are saved;
- **context**: in which are stored the information about a photo belonging to a photographic set or to a group.

For our purposes we selected only a subset of this information. In particular, in our system, we used only photo, user and exif information (fig. 2 shows a simplified logical scheme of our DB). We discarded all the “social” information, which are not useful for our purposes. We discard also tags, which are very noisy, and do not add much useful information to our system. Furthermore we discarded also “place” information as, on Flickr, it does not refer to a specific location but to an area too large for our purposes. Moreover, in our experiments, we observed that, in more than the 40% of the cases, this field refers to an incorrect location. To meet this problem we implemented a geocoding function, exploiting the Google Maps service, which will be discussed in the next subsection. We then defined four views:

- **placeofinterest**: which includes all the information about locations in a given city.

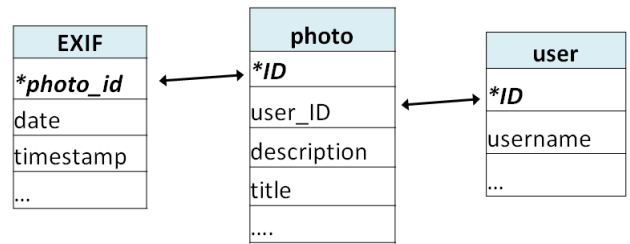


Figure 2. A simplified logical scheme of our DB.

- **photobyuser**: in which are grouped all the photos of a particular user;
- **photobyday**: in which are stored all the image information of a user in a given day;
- **photobylocation**: in which are stored all the images of a user in a given day and for a given location.

The first view is used to build the map of the places of interest. The last three to extract the touristic routes of the Flickr users. Fig.1 shows the scheme of the overall system.

B. Map of the Places of Interest

To fill in the map with the interesting locations of a city, we have to find and locate the places that users visited in that city. At the first step, we retrieve from our DB all the distinct titles of the photos by all the users. Then, for each title, our system calls the geocoding function that queries the Google Maps service, by using the Google Maps API, which returns an URL and the GPS coordinates. The use of the Google Maps service has enabled us to overcome the problems related to multilingualism and synonymy. In fact Google Maps returns the same results independently of the language we use (words as *église*, *chiesa*, *iglesia* or *church* are the same concept in different languages) and of similar meanings (e.g. *building* and *palace*). In this way we transformed the list of titles in a list of Google Maps URLs. Note that different URLs may refer to the same GPS coordinates, if there are different touristic attractions in the same location. Different URLs with the same GPS coordinates are counted as two different places. To validate the retrieved URLs we discard all the locations that are too far away, in terms of their GPS coordinates, from the center of the city, which is a place given by Google Maps. We obviously discard also those titles that do not match with any URLs.

The view *placeofinterest* is then filled with titles of photos, the id of the users who posted the photos with that titles, URLs and GPS coordinates. For each different location, i.e. for each different URL, we count the number of users that visited that place, i.e. those users who posted a photo with a title that matches that URL. Locations are ranked according to the number of visits.

C. Touristic Routes

In this phase we work only with the set of locations extracted and validated in the previous step, i.e. only the photos that can be georeferenced. Since the purpose of our system is the analysis of touristic information, the object of the study are those Flickr users that can be defined as

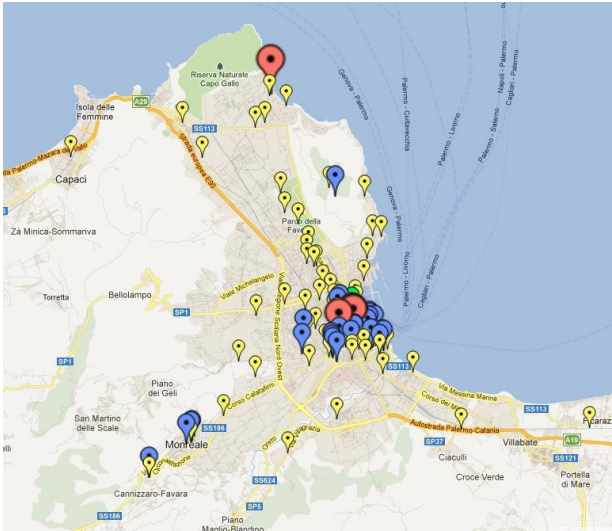


Figure 3. Map of the places of interest in Palermo.

"tourists". To do this, we decided to impose a set of constraints to further filter our data. For a given user (photobyuser) in a given day (photobyday) we select only those photo sets that satisfy the following constraints:

- there must be at least N images taken by the user;
- the time interval between the first and the last picture in the photo set must be at least M hours.

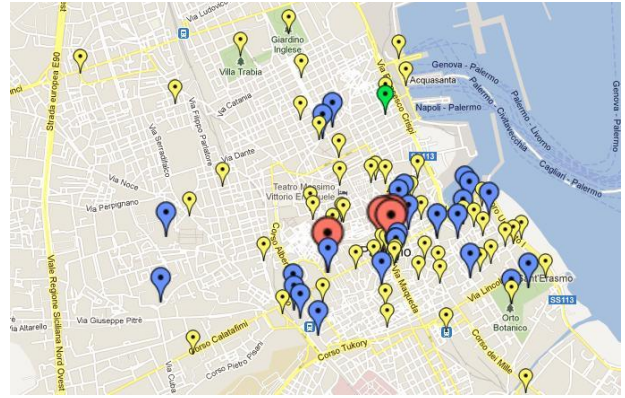
For the first constraint, it is reasonable to assume that a tourist in a day of visit to a city has made a number of photos, as opposed to users, for example, that takes only one photo in a day. The latter type of user does not reflect the category of Flickr users under analysis in this work and therefore, due to this constraint, it is discarded.

In the case of the second constraint, we assume that a tourist, on a day of visiting a city, must have been around at least a certain number of hours, useful to visit places of interest. All photos of that given day which do not satisfy this constraint are discarded.

In a second step, we performed a time analysis onto the places visited by tourists, i.e. the users which satisfied the first set of constraints, to determine if a tourist has actually visited that place. We therefore populate the photobylocation view, with information about the photos taken by a user in a given day, which refer to a specific place. Starting from the previous set of photos, we impose an additional set of constraints:

- There must be at least K pictures of a user for a given day, with the same title;
- The time interval between the first and the last picture, for a given title (place), must be at least L minutes.

With the first constraint, we assume that a tourist, when visiting a site, must have taken at least K pictures with the same title, i.e. of the same subject, otherwise the place is discarded as irrelevant. With the second constraint, we try to estimate the time the tourist has taken to visit a place. It is assumed that a tourist took at least a certain number of minutes to visit a certain place, otherwise the place and their photos are discarded because they are isolated photos.



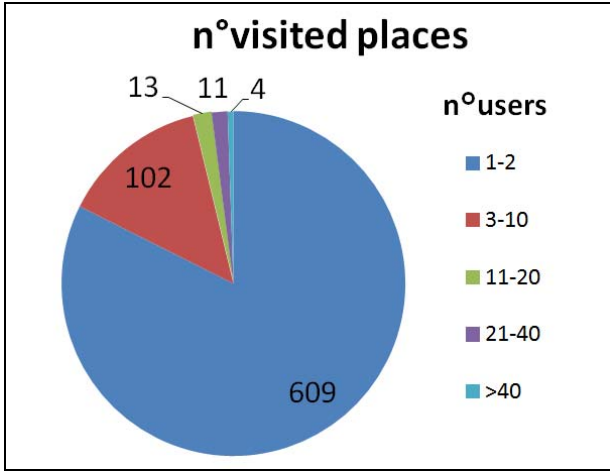


Figure 5. Number of visits. The distribution of the number of visits to the places of interest, i.e. how many places have been visited by a given number of users.

users. Only 28 places were visited by more than 20 users, and these are the ones that actually have some touristic relevance. Many places (609) have been visited by only 1 or 2 users. Most of these sites refer to business activities or personal residences and are not of interest for tourists visiting the city. Within the set of locations that have been visited by more than 2 users, and less than 10, we still measured some false positives (20% ca), i.e. not interesting places or wrong annotations. In case of places that are visited by more than 10 users, no false positives are detected.

With regard to the touristic routes of the users, we repeated our tests varying the constraints described in section 3.3:

- the minimum number of photos taken in one day by a user (PD-number of Photos per Day);
- the minimum number of hours a user must have been around during the day (HD-number of Hours per Day);
- the minimum number of minutes a user takes to visit to a particular place (MP-number of Minutes per Place);
- the minimum number of photos with the same title (PT-number of Photos per Title).

Table 1 summarizes the different sets of parameters that we used in our tests, (each set indicated as a,b,c,d,e,f). We proceeded relaxing one of the constraints, at each test, and we observed the retrieved routes. Table 2 and 3 show the results with the different sets of parameters. As we expected, relaxing the constraints, the number of users that can be considered “tourists”, according to those constraints, increases, and as well the number of retrieved touristic routes. On the other hand, the number of false positives (FP), i.e. wrong or not interesting locations, increases, even if it is acceptable for most of our tests. False positives are typically detected when the title of a photo, or multiple photos, refers to an event (sports, social or personal), rather than a tourist attraction, if this title unfortunately matches with a real physical place on the map. For example, in our tests we found a user who posted some photos about the Palermo Marathon

TABLE I. DIFFERENT SETS OF PARAMETERS

constraints	Parameters set					
	a	b	c	d	e	f
PD	20	10	10	3	3	3
HD	1	1	1	1	0	0
MP	1	1	1	1	1	0
PT	5	5	3	2	2	2

TABLE II. RESULTS WITH THE DIFFERENT SETS OF PARAMETERS

test type	n°users	n°routes	n°FP
a	11	15	0
b	26	31	1
c	42	48	2
d	79	96	8
e	107	139	9
f	116	164	15

(“Maratona di Palermo”) and the Google Maps service has located these photos on the “via Maratona”, which is a road that is located on the outskirts of Palermo, but within the radius defined above. To avoid this type of errors we may decrease the maximum allowed distance parameter of the geocoding function, but that, on the other hand, would exclude some sites that are part of the urban area. It would be therefore necessary a more detailed study on the concept of urban area. Some false positives, however, were detected in the case of incorrect annotations, that matches anyway with some physical places on the map.

When analyzing routes, we can distinguish two types of users: the residents and the tourists. A typical path of a resident user who visits the city is a single-day tour of a single attraction. Nonresident tourists plan their visits trying to see as much interesting places as possible (more than a single place per day) and they often split their visits in more than one day.

In terms of route lengths (tab.3), we observed that, relaxing the constraints, most of the new retrieved routes are made of 1 or 2 places of interest. These are probably related to resident users, as supposed above. If we consider the parameters set with the more restrictive constraints (a), we expect that these results refer to non-resident tourist who visited the city. Fig.6 shows an example of a touristic path of a user visiting the city. Another type of information that can be an interesting object of study is about the number of days a user takes to visit a city (see tab.4). We noted that users which visited the city in 1 or 2 days did not share common or preferred paths, but in most cases these paths are made up of visits to a single place of interest. This behavior may be typical of a resident rather than that of a tourist. Single-day path, that includes more than one place of interest, may instead be related to users from a group of tourists (e.g. a cruise).

Regarding the users which spent more than 2 days visiting the city, about 35% of them visited the city on

TABLE III. NUMBER OF ROUTES WITH A GIVEN LENGTH FOR THE DIFFERENT SETS OF PARAMETERS

n°routes	Parameters set					
route length	a	b	c	d	e	f
1	7	20	29	73	116	123
2	3	6	10	11	11	15
3	1	1	2	4	4	10
4	2	2	4	3	3	4
5	0	0	1	2	2	1
6	1	1	0	0	0	1
7	1	1	0	1	1	0
8	0	0	1	0	0	0
9	0	0	1	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	1	1	1
13	0	0	0	1	1	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	0	0	0	0	0	1

TABLE IV. NUMBER OF USERS THAT VISITED THE CITY FOR A GIVEN NUMBER OF DAYS FOR THE DIFFERENT SETS OF PARAMETERS.

users	days of visit				
test type	1	2	3	4	5
a	10	0	0	0	1
b	24	1	0	0	1
c	39	2	0	0	1
d	65	11	2	0	1
e	84	18	2	2	1
f	89	15	6	3	3

consecutive days, while 65% of these users have not taken pictures in the same period (in some cases even in different years). It can be assumed that the typical behavior of a tourist visiting the city is represented by the first type of users (who make more consecutive days of visit). For the second type of user this behavior is similar to that of a resident person which decide to visit some interesting places of the city.

V. CONCLUSIONS

In this paper we presented a GIR system for extracting and analyzing information from metadata associated with touristic photos. The purpose of this work was to obtain statistical information about the touristic visits of the users of Flickr and, from this, also derive information about major touristic attractions of a city. We have chosen as a case study the city of Palermo, but our system can be used to analyze any location in the world, if it is possible to obtain a statistically significant number of photos of that location from Flickr users.

The main problem faced during our work was the lack of accurate and reliable annotations by the Flickr users. For this reason our system has to discard a large number of images that do not contain useful information for the georeferencing. In addition, metadata about geospatial

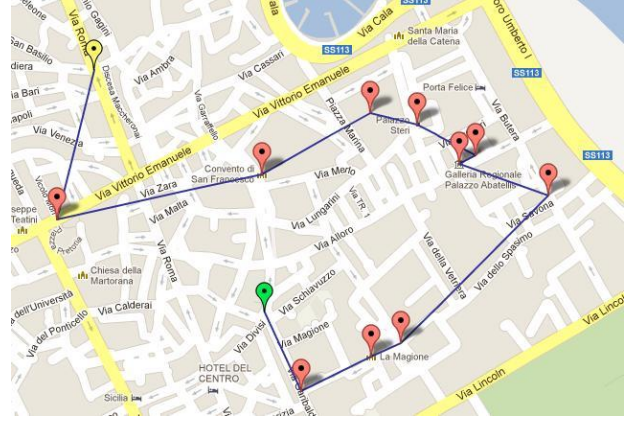


Figure 6. Figure 6 An example of a touristic route around the center of the city. Green marker is the starting point, yellow marker is the end point.

information provided by Flickr are too generic and often incorrect. For this reason it was necessary to exploit another tool, Google Maps, to geolocate user photos. In this way we use photo titles to search for GPS coordinates by querying Google Maps. This service does not suffer from the problems of multilingualism and synonymy, then photos of the same place, which have different titles (as written in different languages or using similar words), can be georeferenced into the same location.

A possible improvement of our system could be to enrich the Flickr metadata, for example giving a score to the tags, that are discarded in our work, to take only the most reliable ones. Our system can be also improved with the use of specific dictionaries, in which are contained all the information about the places of a city, in order to validate the identification of the places of interest.

Another critical issue is the execution time to extract and analyze the data. To reduce this problem we downloaded data of interest into a local database, eliminating the problem of the interaction with a the remote server, but this phase required a considerable processing time. Nevertheless, some of the tests still took about a couple of hours of processing, since the amount of data is high and information is processed sequentially. The simplest solution would be to exploit the parallel computation for both the DB population and the data processing steps.

The proposed system can be a very useful tool for any organization that work in tourism, who want to get more information about the most preferred attractions of a city, and to study the favorite paths of the tourists in that city.

In future works we plan to use the extracted information to suggest new touristic paths to the users, depending on their constraints (visiting time, starting place, specific interests, etc.)

REFERENCES

- [1] R. R. Larson, "Geographic information retrieval and spatial browsing," in *Geographic Information Systems Patrons Maps and Spatial Information*, pp 81-124.

- [2] N. Sawant, J. Li, J. Wang, "Automatic image semantic interpretation using social action and tagging data," *Multimedia Tools Appl.*, vol. 51 pp. 213-246, .
- [3] I. Simon, N. Snavely, S. M. Seitz, "Scene Summarization for Online Image Collections", *IEEE 11th International Conference on In Computer Vision*, 2007, pp.1-8.
- [4] S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, R. Szeliski, "Building Rome in a day," *Computer Vision, 2009 IEEE 12th International Conference on* , vol., no., pp.72-79, Sept. 29 2009-Oct. 2 2009.
- [5] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. "Generating summaries and visualization for large collections of geo-referenced photograph". In *Proceedings of the 8th ACM international workshop on Multimedia information retrieval (MIR '06)*, pp. 89-98.
- [6] S. Ahern, M. Naaman, R. Nair, and J. H. Yang., "World explorer: visualizing aggregate data from unstructured text in geo-referenced collections," In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries (JCDL '07)*, pp. 1-10.
- [7] L. Kennedy, M. Naaman, S. Ahern, R. Nair, and T. Rattenbury, "How flickr helps us make sense of the world: context and content in community-contributed media collections," In *Proceedings of the 15th international conference on Multimedia (MULTIMEDIA '07)*, pp.631-640.
- [8] W. C. Chen, A. Battestini, N. Gelfand, and V. Setlur, "Visual summaries of popular landmarks from community photo collections," In *Proceedings of the 17th ACM international conference on Multimedia (MM '09)*, pp.789-792.
- [9] T. Rattenbury, N. Good, and M. Naaman, "Towards automatic extraction of event and place semantics from flickr tags," In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '07)* pp. 103-110.
- [10] A. Popescu and G. Grefenstette, "Deducing trip related information from flickr," In *Proceedings of the 18th international conference on World wide web (WWW '09)*, pp. 1183-1184.
- [11] Y. Ren, M. Yu, X.J. Wang, L., Zhang, and W. Y. Ma. "Diversifying landmark image search results by learning interested views from community photos," In *Proceedings of the 19th international conference on World wide web (WWW '10)*, pp. 1289-1292.
- [12] A. J. Cheng, Y. Y. Chen, Y. T. Huang, W. H. Hsu, and H. Y. M. Liao, "Personalized travel recommendation by mining people attributes from community-contributed photos", In *Proceedings of the 19th ACM international conference on Multimedia (MM '11)*, 83-92.